Title: Ten simple rules for working with high resolution remote sensing data

Maxwell B. Joseph1, Anna I. Spiers1,2, Michael J. Koontz1, Nayani Ilangakoon1, Kylen Solvik1,3, Nathan Quarderer1, Joe McGlinchy1, Victoria M. Scholl1,3, Lise St. Denis1, Chelsea Nagy1,3, Anna Braswell4, Matthew W. Rossi1, Lauren Herwehe1,3, Leah Wasser1,3, Megan E. Cattau5, Virginia Iglesias1, Adam L. Mahood1,3,\*, Fangfang Yao3, Stefan Leyk1,3,6, Jennifer K. Balch1,3,

1 Department of Geography, University of Colorado Boulder, Boulder, CO, USA

2 Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder, CO, USA

3 Earth Lab, University of Colorado, Boulder, CO, USA

4 School of Forest Resources and Conservation, University of Florida, Gainesville, FL, USA

5 Department of Human-Environment Systems, Boise State University, Boise, ID, USA

6 Institute of Behavioral Science, University of Colorado Boulder, Boulder, CO, USA

\* Corresponding author: [admahood@gmail.com](mailto:admahood@gmail.com)

# **Abstract**

Researchers in Earth and environmental science can extract incredible value from high resolution remote sensing data, but these data can be hard to use. Pain free use requires skills from remote sensing and the data sciences that are seldom taught together. In practice, many researchers teach themselves how to use high resolution remote sensing data with ad hoc trial and error processes, often resulting in wasted effort and resources. Here we outline ten “rules” with examples from Earth and environmental science to help applied researchers work more effectively with high resolution data.

# Introduction

The data revolution brings a deluge of Earth observations from numerous and diverse sensors. Many of these data are collected remotely: from space, the air, or underwater, and are of increasingly high resolution, providing detailed spatial, temporal, radiometric, and/or spectral information (Figure 1). Earth and environmental scientists increasingly use high resolution remote sensing data, but learning how to do so can be difficult. In this article, we outline 10 simple rules to help applied Earth and environmental researchers use high resolution remote sensing data.

High resolution data allow us to answer persistent science questions in different ways, and to ask new questions. For instance, the Shuttle Radar Topography Mission generated a near-global digital elevation model (DEM) at 30m resolution ([Farr and Kobrick 2000](#ref-farr2000shuttle)) which provided new insights into hydrography ([Lehner, Verdin, and Jarvis 2008](#ref-lehner2008new)), cryology ([Surazakov and Aizen 2006](#ref-surazakov2006estimating)), vegetation remote sensing ([Simard et al. 2006](#ref-simard2006mapping)), climate change-induced coastal flood risk ([McGranahan, Balk, and Anderson 2007](#ref-mcgranahan2007rising)) and more. But, what defines “high resolution” changes over time, and a 30m DEM is considered quaint today, relative to sub-meter topography data that are increasingly available and yielding new insights [Thatcher, Lukas, and Stoker](#ref-thatcher20203d) ([2020](#ref-thatcher20203d)). For instance, a novel integration of SRTM with higher resolution lidar-derived elevation data tripled the estimate of the number of people at risk worldwide from coastal flooding in the next century ([Kulp and Strauss 2019](#ref-kulp2019new)). Although high resolution data are valuable, they are not always easy to use and can be of limited benefit in some cases.

Pain-free use of high resolution data requires remote sensing and data science skills that not all applied scientists have ([Hampton et al. 2017](#ref-hampton2017skills)). High resolution data can be voluminous, complex, and noisy, requiring systematic data management, workflow management, and data processing skills as well as in-depth uncertainty assessments. Further, high resolution remote sensing data are often integrated with other sources of information (e.g., ground truth data or other environmental data), which brings additional challenges associated with data harmonization, reconciliation, and uncertainty propagation ([Zipkin et al. 2021](#ref-zipkin2021addressing)). In practice, learning how to use high resolution data is often an ad hoc trial and error process. The resulting bespoke approaches that applied researchers develop can be inconsistent, inefficient, and challenging to implement, reproduce, or extend ([Wyngaard et al. 2019](#ref-wyngaard2019emergent)).

Here we outline a set of “rules” to provide a foundation that applied researchers can build upon to work effectively with high resolution data. We focus on examples in Earth and environmental science, but the ideas apply to other disciplines.

# 1. Know your question

While high resolution data can be exciting to work with, prioritize your science question. Know why the question matters, and anticipate potential answers. Write the question down, so that you can return to it throughout the lifespan of a project. Use the question to figure out when the project is done, i.e., at what point you will have answered the question. A clear question can also help with understanding data requirements including spatial, temporal, radiometric, and spectral resolutions and extents (see Know your data).

Suppose you are asking a question about local plant population dynamics. You may need high resolution data to identify individual plants in a small region ([Koontz et al. 2021](#ref-koontz2021cross)). In contrast, a question about vegetation and large-scale wildebeest migration may require coarser resolution vegetation index data over a large geographic area ([Musiega, Sanga-Ngoie, and Fukuyama 2006](#ref-musiega2006framework)). Finally, even high resolution data may be sampled from a larger population of data. If your science question requires inference about this larger population, it is important to understand whether the available sample of data permits inference. For example, spatial bias in data availability can lead to unrepresentative samples, complicating large-scale inference ([Metcalfe et al. 2018](#ref-metcalfe2018patchy)).

Clearly state a science question, and make it a good one ([Alon 2009](#ref-alon2009choose)). Know the scope and key attributes of what you are measuring, including scale, resolution, and level of organization (e.g., individual, community, ecosystems, landscape), to choose the most appropriate data. Consider sample representativeness and mismatch between the phenomenon scale and the analytical scale you are using. Use domain expertise on your research team to identify potential challenges at the interface of your question and available data.

Identify the frontiers of research in your field and state a question. A well-posed question points to data requirements and a clear end point.

# 2. Know your data

In addition to defining your science question, you need to know your data. This includes knowing whether the available data are fit for the intended use, given their assumptions, biases, strengths and limitations. The concept of fitness for use of a data product is useful for assessing the relative data quality ([Tayi and Ballou 1998](#ref-tayi1998examining)), and appropriateness of a given dataset for an intended purpose ([Agumya and Hunter 1999](#ref-agumya1999risk); [Bruin, Bregt, and Ven 2001](#ref-bruin2001assessing); [Devillers et al. 2007](#ref-devillers2007towards)). Key considerations include: can data measure the phenomena of interest, and how does the scale of the data relate to the scale of the phenomena (see Know your question).

For example, 30m Landsat pixels cannot tell us much about when individual trees turn green. Here, an uncrewed aerial system (UAS) would be more fitting, as it can collect sub-meter data with a customizable revisit time for local-scale analyses ([Anderson and Gaston 2013](#ref-anderson2013lightweight)). Even with a UAS, particular sensors have tradeoffs and limitations to consider. For instance, two technologies are often compared in forest mapping applications: Structure from Motion (SfM) photogrammetry and light detection and ranging (lidar). SfM uses multiple images to construct 3D models, is less expensive, and has well- established processing workflows ([Westoby et al. 2012](#ref-westoby2012structure)). Science-grade lidar systems are more accurate and more expensive. Investing in the resources for science-grade lidar data collection and processing has proved to be worthwhile in forests with dense canopies ([Lefsky et al. 2002](#ref-lefsky2002lidar)). In other cases, SfM is an adequate low-cost alternative ([Wallace et al. 2016](#ref-wallace2016assessment)), especially in developing countries where funds may be limited ([Mlambo et al. 2017](#ref-mlambo2017structure)).

To start, you should 1) explore how and why your data were collected and how they were processed (raw, secondary or modeled data); 2) understand what the data measure and 3) understand potential errors, biases, and uncertainties within your data. This includes spatial data quality components such as positional, attribute, and semantic accuracy, completeness, logical consistency, etc. ([Guptill and Morrison 2013](#ref-guptill2013elements)). You can build this understanding by reading original descriptions of data products in the peer reviewed literature, data product user guides, algorithm theoretical basis documents, product specification reports, and digging into metadata. It can also be helpful to work with outside experts to better understand fitness for use. If you are the expert, you may need to carry out your own assessment using reference data. Finally, if no one data source suffices, consider whether data fusion or integration is possible ([Schmitt and Zhu 2016](#ref-schmitt2016data)). This can be complicated however by a need for resampling, aggregation, reprojection, or interpolation resulting in complex uncertainty propagation. Such modifications, which are often ignored but can affect inference, have to be addressed.

Know your data’s characteristics, strengths, and weaknesses, so you can determine whether it is appropriate. Understand data uncertainty, uncertainty propagation, and the implications for your application.

# 3. Don’t use high resolution data

If coarse resolution data suffice, avoiding high resolution data can reduce time requirements, complexity, and costs (both computational and monetary). Deciding whether to use high resolution data requires a clear vision of how different data products align with the goals of a project, and knowledge about the effort required to use alternative data products. The decision-making process should be based on principles of scale sensitivity and efficiency.

Coarse spatial resolution data may work well for phenomena operating at regional to continental scales, depending on the project goals. For example, volcanic ash plumes are detectable with kilometer-sized pixels, and low spatial/high temporal resolution data from geostationary satellites might suffice when measuring global ash transport ([Woods, Holasek, and Self 1995](#ref-woods1995wind)). To measure ash deposition on buildings or vehicles, a higher spatial resolution data product would be necessary.

High resolution data should be weighed against lower resolution alternatives, guided by science needs (see Know your question), cost/benefit analysis, ethical considerations (see Do no harm) and practical constraints. If the decision is difficult to make, consider starting with lower resolution data to better understand the need for finer granularity, or a sample of fine-resolution (often large volume) data to be able to run models or processes efficiently.

High resolution data are invaluable when needed, but using high resolution data requires additional time, effort, and computational resources. If coarser data meet your science needs, maybe don’t use high resolution data.

# 4. Know when to innovate

Can we use existing data or approaches to answer science questions, or do we need to innovate with new approaches or new high resolution data products? Innovation may be costly (see Survey the computing landscape), and the answer to this question depends on the expected return on your investment. Sometimes, using an existing dataset or method may be a better option, e.g., when existing methods are good enough and your primary goal is not methods development (see Keep your focus). Faced with the options of using new high resolution data with old methods, or developing new methodology tailored to high resolution data, how can you decide whether or not to innovate?

Sometimes existing approaches work just fine. For example, [Wyder et al.](#ref-wyder2019autonomous) ([2019](#ref-wyder2019autonomous))], tracked moving objects in real-time using drones with a neural network-based object detector (You Only Look Once, or YOLO ([Redmon et al. 2016](#ref-redmon2016you))). While this algorithm does not have the best detection accuracy among similar, more computationally intensive algorithms (e.g., deeper neural networks, or architectures that explicitly model sequences of images), YOLO is computationally efficient, allowing for high frame rate object detection with limited computing power. In other cases, methodological innovation can overcome data limitations. For example, although high point density lidar data contain information about individual tree canopies, training an object detector to identify individual trees is difficult because of a lack of training data (hand-labeled bounding boxes around individual canopies). This issue can be addressed with weakly supervised learning, where models are pre-trained using many poor-quality bounding boxes that are cheap to generate, and then fine-tuned using a much smaller dataset of high-quality bounding boxes [weinstein2020cross].

Performing a thorough literature review will ensure that your research is well-informed by the progress already made in your field ([Boote and Beile 2005](#ref-boote2005scholars)). It may not be appropriate to use traditional approaches with data at higher resolutions; this is the time to consider unique opportunities that were not possible before. Look beyond the boundaries of your field or discipline for new ideas, approaches, and perspectives ([Shaman et al. 2013](#ref-shaman2013fostering)), but try to Keep your focus. The cost of innovation needs to be weighed against the value of the information gained. Consider whether energy invested in developing a method will lower research or technical debt later ([Olah and Carter 2017](#ref-olah2017research)). If you do choose to innovate, Show your work and create open workflows to ensure that your effort is also accessible to the community.

Weigh the pros and cons of innovation for your particular project. Don’t reinvent the wheel.

# 5. Keep your focus

High-resolution datasets are information-rich, with many potentially exciting science applications to explore. This supports new discoveries (see Allow for the unexpected), and methods (see Know when to innovate), but it can be easy to get distracted from your science question, lost in tangential gee whiz inquiries. While scope creep is not always bad, it is important keep in mind: is your main goal to develop a new method or to investigate a particular phenomenon? You might need to do both, but one should be the focus and the other should serve in a supporting role during the research process.

For example, you start a project to detect individual trees from high-resolution hyperspectral imagery. In your initial data exploration, you realize that you might be able to distinguish individual tree species based on their spectral signatures. You could easily spend weeks or months exploring species classification, only to realize you have made little progress on your original problem: identifying individual trees regardless of species. As another example, let us say that you did not get distracted by that tangent and you developed a tree classification that performs well in 95 percent of your study region, but in a specific corner of the forest it performs very poorly. You face a decision: do you try a new, more complex method on the whole region, or do you stop and simply report the poor performance as a model caveat?

Early-stage planning can help you keep focus. Once you have chosen your question (see Know your question), carefully define the scope of your project and aim to start small (see Start small). Straying outside that scope and allowing for a tangential inquiry can be helpful, but it is important to have a strategy from the outset to decide how much time and effort you can spare for tangential inquiries. In the end, it may be better to save some things for future projects, and you might choose to keep a parking lot of ideas you do not end up exploring fully and so that you can return to them later. Your results do not have to be perfect and all-encompassing to have meaningful impact. Science most often advances in small steps. A case study or a negative result is still a valuable contribution.

Define and (mostly) stick to the scope of your project, revisiting it as you work. Do not let the perfect be the enemy of the good.

# 6. Survey the computing landscape

High resolution data can be time- and resource-intensive, so it is important to survey the software landscape to identify existing tools that will help build efficient workflows. Understanding your hardware, memory, and CPU requirements will speed up the iterative process of writing code, troubleshooting bugs, and developing analyses. Know which computing platforms meet your requirements, whether it be in the cloud, a high performance computing cluster, or a local workstation.

Often, the data used define the software needed. For example, National Ecological Observatory Network (NEON) aerial hyperspectral imagery have 426 spectral bands spanning the visible to shortwave infrared wavelengths of the electromagnetic spectrum ([Kampe et al. 2010](#ref-kampe2010neon)). One file may cover 7.5 km and can be on the order of 2.5 GB compressed in the HDF5 (hierarchical data) format. When loaded into memory as a numerical array it can require close to 26 GB of memory (e.g., a 6307x1239x426 floating point array). Many personal computers can not load the data in memory. However, the file format of the data supports both compression and slicing operations with open source Python tools such as Xarray and Dask, allowing the data to be referenced and loaded only when computation is required, and distributing computations across multiple processors [Hoyer and Hamman](#ref-hoyer2017xarray) ([2017](#ref-hoyer2017xarray)). These tools can enable analyses that would be impossible otherwise.

Research whether there are existing software tools that have already been created and optimized to load and process your data. For instance, the neonhs R package enables efficient access to NEON hyperspectral imagery ([Joseph 2021](#ref-max_joseph_2021_4641288)). Packages that are stable and actively maintained and/or supported by rOpenSci and pyOpenSci can provide a good starting point ([Boettiger et al. 2015](#ref-boettiger2015building)). Seek tools from other disciplines that might prove useful (see Know when to innovate). For instance, the cloth simulator filter algorithm for classifying “ground” versus “not ground” in lidar or SfM photogrammetry point clouds is both accurate and efficient for this purpose, though it was originally developed for efficiently mimicking the movement of fabric in video games ([Zhang et al. 2016](#ref-zhang2016easy)).

You would not start building a house without knowing what tools you need. Invest time early in a project to understand which tools will help get your work done.

# 7. Start small

Developing a workflow is an iterative process. Given the large volume of high resolution data, iteration costs can be time-intensive and computationally expensive. Start small with subsets of data and simpler models to enable rapid iteration and experimentation. When working with data subsets, it is useful to identify minimum iterable units: the smallest units in the data that can be treated independently for computation. Test your workflow on a small fraction of those iterable units before applying it to the entire dataset to increase your workflow efficiency.

For example, in a study of wet-dry dynamics of 71,842 playa lakes on the Great Plains, monthly Landsat-derived water history data were extracted for us with a machine learning model ([Solvik et al. 2020](#ref-solvik2020predicting)). Data extraction was prototyped on a few playa lakes (the minimum iterable units), until an efficient method was developed. Similarly, initial models focused on training a time series model using data from just a few playa lakes. These early modeling steps can ensure that your workflow is functional at low cost. Similarly, a study mapping the microtopography of ice wedges in Alaska over 1200 square km landscape using high resolution lidar data dealt with the enormous data volume, by first training a convolutional neural network model using a small, representative subset of data on a laptop which took 30 minutes ([Abolt and Young 2020](#ref-abolt2020high)). Then a model was trained on the entire dataset in parallel on a cluster.

Start by applying the simplest tractable model over a small representative sample of minimum iterable units. Iterative experimentation with high volume, high resolution data at scale is a recipe for wasted time and resources. Ideally, there should be rapid feedback when trying something new that helps to guide your work. Knowing whether an approach works within minutes or hours will be more efficient than waiting days or weeks to realize that your code or model is broken.

Start small with a prototype, model, or data subset to maximize efficiency, identify errors, and test workflows with a low-cost representative subset of the data.

# 8. Allow for the unexpected

High resolution images can provide clearer pictures of objects. The additional detail from high resolution data may allow novel or unexpected information to emerge about your system. While we do recommend focusing on a science question, we would be remiss not to acknowledge that high resolution data support unexpected scientific discoveries. This is especially true for high resolution data that are early-stage, cutting edge, or being used in a new area.

For example, high resolution lidar has uncovered previously undescribed archaeological sites ([Bewley, Crutchley, and Shell 2005](#ref-bewley2005new)) and active faults ([Hunter et al. 2011](#ref-hunter2011lidar)). High resolution lidar data of the ground surface and vegetation canopy structure have also revealed complex interactions between soils, termites, and hydrology that explain the spatial distributions of plants and termite mounds in savanna ecosystems ([Levick et al. 2010](#ref-levick2010regional)). Carbon stock estimation is another such example. Measuring carbon stocks and their response to disturbance has historically been limited to regional extents ([Asner et al. 2014](#ref-asner2014targeted)), but with new spaceborne missions (e.g., Global Ecosystem Dynamics Investigation ([Dubayah et al. 2020](#ref-dubayah2020global))), we can scale these approaches to the continental scale.

High resolution remote sensing has the potential of revealing new phenomena, features, and processes. As users of such data, this can be a unique opportunity for discovery. However, not everything that is unexpected leads to useful insights. Pursuing such lines of inquiry could be rewarding, but carries a risk of distracting you from your original goals.

If you see something unexpected or novel with high resolution data, take note. But, consider finishing your plate (answering your science question) before eating dessert.

# 9. Do no harm

High resolution data carry risks for unintended or malicious use. The demarcation of municipal and property boundaries, risk and hazard assessment, real-time surveillance, and public health monitoring are all areas that benefit from data collected at fine spatial and/or temporal scales. And while those who gather and distribute high resolution mapping data may have everyone’s best interest in mind, there is potential harm associated with collection and redistribution of high resolution data. Care needs to be taken to ensure ethical data use, but who decides what is ethical? Such issues become even more prominent as data from multiple sources become synthesized to identify events, processes, or phenomena that could not otherwise be detected using a single data source alone, potentially resulting in unintended violations of privacy.

For example, uncrewed aerial vehicles (UAV) can track the movement of displaced populations ([Berman et al. 2018](#ref-berman2018ethical)). High resolution satellite imagery can identify evidence of war crimes, or track environmental impacts associated with mining and deforestation ([Harris 2013](#ref-harris2013reflections)). While these applications have the potential to benefit certain parties, some see these sorts of observations as a threat to safety and wellbeing, putting those who are already vulnerable at further risk of surveillance by bad actors ([Wang 2019](#ref-wang2019success)). Other examples include sharing locations of sacred and historic sites of burial or worship, nesting sites of endangered species, and the movement of military assets.

It is critical to consider harm that could result from use of high resolution data. There may be moral challenges associated with providing sub-meter resolution imagery at a global scale to anyone with a standard internet connection. Practitioners should take this into consideration when collecting, storing, distributing, and using such data. We suggest that effort be made to protect the privacy and confidentiality of stakeholders or third parties, and to obtain consent whenever possible prior to data collection or use. If the same questions can be answered without high resolution data, consider using coarser data (see Don’t use high resolution data). Ask yourself how storing or sharing data could compromise stakeholder privacy. Think about what could happen if your data or analysis fell into the wrong hands. If it could do harm, consider whether to proceed and how to mitigate harm.

Identify risks associated with data collection, storage, and dissemination. Steps to mitigate against ethical conflicts include measures to acquire consent, protect privacy, and provide transparency.

# 10. Show your work

In the interest of open reproducible science, it is important to “show your work” ([Munafò et al. 2017](#ref-munafo2017manifesto)). By showing your work, you can increase the transparency, reproducibility, and usability of your science ([Hampton et al. 2015](#ref-hampton2015tao)). This makes your life easier if you need to re-run analyses later on, and more broadly supports scientific discovery and innovation for entire communities of end users ([Lowndes et al. 2017](#ref-lowndes2017our)).

In some applications, there is tension between accessible open research and the practical reality of working with high resolution data which may involve expensive commercial software or data or ethical concerns (see Do no harm). For example, Agisoft provides a robust closed source platform to create 3D models from drone imagery. For many researchers, commercial software may be cheaper and more accessible than developing an open source alternative ([Li et al. 2016](#ref-quan2016construction)). Google Earth Engine similarly is proprietary but provides unprecedented access to many high resolution data products that would otherwise be out of reach for many researchers. These trade-offs can also arise with data, e.g., commercial satellite imagery may be expensive but necessary ([McGlinchy et al. 2019](#ref-mcglinchy2019application)). In these cases, you can work toward reproducibility by: 1) disclosing all data and steps used in a workflow, 2) reporting all algorithms (with citations) and settings used in a data pipeline, and 3) if possible, modularizing your workflow so that other tools and/or data can be substituted in the future.

Regardless of the application, the data, software and workflows that you generate can be guided by the principles of findability, accessibility, interoperability, and reusability ([Wilkinson et al. 2016](#ref-wilkinson2016fair)). These principles can be translated to a variety of specific actions such as providing open access to your original and derived data products, documenting and releasing software, recording and reporting metadata, releasing end-to-end workflows or data pipelines, and building research compendia around publications ([Gray and Marwick 2019](#ref-gray2019truth)).

Challenges arise when doing open reproducible research with high resolution remote sensing data, but navigating this space is critical to make your work more useful. Show your work, even if you are using proprietary data or software.

# Conclusion

These rules represent practical advice for working with high resolution remote sensing data as an applied researcher in the Earth and environmental science data revolution ([Kitchin 2014](#ref-kitchin2014data)). Although the definition of “high resolution” is fluid, and future remote sensing data might provide unforeseen advances in spatial, temporal, spectral, and radiometric resolution, we expect that these general principles will hold as future generations of remote sensing data emerges over the coming decades. Eventually, it would be nice if the training for applied scientists provided all of the data science and remote sensing skills required to work with high resolution remote sensing data, and this paper would no longer be necessary. In the meantime, we hope that these simple rules provide some useful guidance.

# Author Contributions

MBJ had the initial conception of the project, organized the collaborative working sessions, co-authored two rules, and drafted the introduction and conclusion. MWR and JM co-authored one and a half rules, helped with overall revisions, and created Figure 1. ALM, AIS, VMS, NI, LAS, NQ, MEC, KS, LH, AB, RCN, and VI co-authored two rules and helped with overall revisions. LW, FY, MJK and SL co-authored one rule and helped with overall revisions. JKB co-authored one rule, organized and funded the working group. MWR and ALM cracked both bad and good jokes, respectively. Author order is randomized after MBJ.

# References

Abolt, Charles J, and Michael H Young. 2020. “High-Resolution Mapping of Spatial Heterogeneity in Ice Wedge Polygon Geomorphology Near Prudhoe Bay, Alaska.” *Scientific Data* 7 (1): 1–7.

Agumya, Aggrey, and Gary J Hunter. 1999. “A Risk-Based Approach to Assessing the Fitness for Use of Spatial Data.” *URISA Journal* 11 (1): 33–44.

Alon, Uri. 2009. “How to Choose a Good Scientific Problem.” *Molecular Cell* 35 (6): 726–28.

Anderson, Karen, and Kevin J Gaston. 2013. “Lightweight Unmanned Aerial Vehicles Will Revolutionize Spatial Ecology.” *Frontiers in Ecology and the Environment* 11 (3): 138–46.

Asner, Gregory P, David E Knapp, Roberta E Martin, Raul Tupayachi, Christopher B Anderson, Joseph Mascaro, Felipe Sinca, et al. 2014. “Targeted Carbon Conservation at National Scales with High-Resolution Monitoring.” *Proceedings of the National Academy of Sciences* 111 (47): E5016–22.

Berman, Gabrielle, Sara de la Rosa, Tanya Accone, and others. 2018. “Ethical Considerations When Using Geospatial Technologies for Evidence Generation.” *Innocenti Discussion Papers*.

Bewley, Robert H, Simon P Crutchley, and Colin A Shell. 2005. “New Light on an Ancient Landscape: Lidar Survey in the Stonehenge World Heritage Site.” *Antiquity* 79 (305).

Boettiger, Carl, Scott Chamberlain, Edmund Hart, and Karthik Ram. 2015. “Building Software, Building Community: Lessons from the rOpenSci Project.” *Journal of Open Research Software* 3 (1).

Boote, David N, and Penny Beile. 2005. “Scholars Before Researchers: On the Centrality of the Dissertation Literature Review in Research Preparation.” *Educational Researcher* 34 (6): 3–15.

Bruin, Sytze de, Arnold Bregt, and Marc van de Ven. 2001. “Assessing Fitness for Use: The Expected Value of Spatial Data Sets.” *International Journal of Geographical Information Science* 15 (5): 457–71.

Devillers, Rodolphe, Yvan Bédard, Robert Jeansoulin, and Bernard Moulin. 2007. “Towards Spatial Data Quality Information Analysis Tools for Experts Assessing the Fitness for Use of Spatial Data.” *International Journal of Geographical Information Science* 21 (3): 261–82.

Dubayah, Ralph, James Bryan Blair, Scott Goetz, Lola Fatoyinbo, Matthew Hansen, Sean Healey, Michelle Hofton, et al. 2020. “The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth’s Forests and Topography.” *Science of Remote Sensing* 1: 100002.

Farr, Tom G, and Mike Kobrick. 2000. “Shuttle Radar Topography Mission Produces a Wealth of Data.” *Eos, Transactions American Geophysical Union* 81 (48): 583–85.

Gray, Charles T, and Ben Marwick. 2019. “Truth, Proof, and Reproducibility: There’s No Counter-Attack for the Codeless.” In *Research School on Statistics and Data Science*, 111–29. Springer.

Guptill, Stephen C, and Joel L Morrison. 2013. *Elements of Spatial Data Quality*. Elsevier.

Hampton, Stephanie E, Sean S Anderson, Sarah C Bagby, Corinna Gries, Xueying Han, Edmund M Hart, Matthew B Jones, et al. 2015. “The Tao of Open Science for Ecology.” *Ecosphere* 6 (7): 1–13.

Hampton, Stephanie E, Matthew B Jones, Leah A Wasser, Mark P Schildhauer, Sarah R Supp, Julien Brun, Rebecca R Hernandez, et al. 2017. “Skills and Knowledge for Data-Intensive Environmental Research.” *BioScience* 67 (6): 546–57.

Harris, Ray. 2013. “Reflections on the Value of Ethics in Relation to Earth Observation.” *International Journal of Remote Sensing* 34 (4): 1207–19.

Hoyer, Stephan, and Joe Hamman. 2017. “Xarray: ND Labeled Arrays and Datasets in Python.” *Journal of Open Research Software* 5 (1).

Hunter, LE, JF Howle, RS Rose, and GW Bawden. 2011. “LiDAR-Assisted Identification of an Active Fault Near Truckee, California.” *Bulletin of the Seismological Society of America* 101 (3): 1162–81.

Joseph, Maxwell B. 2021. “Earthlab/Neonhs: V0.0.1.” Earth Lab, University of Colorado Boulder; Zenodo. <https://doi.org/10.5281/zenodo.4641288>.

Kampe, Thomas U, Brian Robert Johnson, Michele A Kuester, and Michael Keller. 2010. “NEON: The First Continental-Scale Ecological Observatory with Airborne Remote Sensing of Vegetation Canopy Biochemistry and Structure.” *Journal of Applied Remote Sensing* 4 (1): 043510.

Kitchin, Rob. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Sage.

Koontz, Michael J, Andrew M Latimer, Leif A Mortenson, Christopher J Fettig, and Malcolm P North. 2021. “Cross-Scale Interaction of Host Tree Size and Climatic Water Deficit Governs Bark Beetle-Induced Tree Mortality.” *Nature Communications* 12 (1): 1–13.

Kruse, Fred A, William M Baugh, and Sandra L Perry. 2015. “Validation of DigitalGlobe WorldView-3 Earth Imaging Satellite Shortwave Infrared Bands for Mineral Mapping.” *Journal of Applied Remote Sensing* 9 (1): 096044.

Kulp, Scott A, and Benjamin H Strauss. 2019. “New Elevation Data Triple Estimates of Global Vulnerability to Sea-Level Rise and Coastal Flooding.” *Nature Communications* 10 (1): 1–12.

Lefsky, Michael A, Warren B Cohen, Geoffrey G Parker, and David J Harding. 2002. “Lidar Remote Sensing for Ecosystem Studies: Lidar, an Emerging Remote Sensing Technology That Directly Measures the Three-Dimensional Distribution of Plant Canopies, Can Accurately Estimate Vegetation Structural Attributes and Should Be of Particular Interest to Forest, Landscape, and Global Ecologists.” *BioScience* 52 (1): 19–30.

Lehner, Bernhard, Kristine Verdin, and Andy Jarvis. 2008. “New Global Hydrography Derived from Spaceborne Elevation Data.” *Eos, Transactions American Geophysical Union* 89 (10): 93–94.

Levick, Shaun R, Gregory P Asner, Oliver A Chadwick, Lesego M Khomo, Kevin H Rogers, Anthony S Hartshorn, Ty Kennedy-Bowdoin, and David E Knapp. 2010. “Regional Insight into Savanna Hydrogeomorphology from Termite Mounds.” *Nature Communications* 1 (1): 1–7.

Li, Xiu quan, Zhu an Chen, Li ting Zhang, and Dan Jia. 2016. “Construction and Accuracy Test of a 3d Model of Non-Metric Camera Images Using Agisoft PhotoScan.” *Procedia Environmental Sciences* 36: 184–90.

Lowndes, Julia S Stewart, Benjamin D Best, Courtney Scarborough, Jamie C Afflerbach, Melanie R Frazier, Casey C O’Hara, Ning Jiang, and Benjamin S Halpern. 2017. “Our Path to Better Science in Less Time Using Open Data Science Tools.” *Nature Ecology & Evolution* 1 (6): 1–7.

McGlinchy, Joe, Brian Johnson, Brian Muller, Maxwell Joseph, and Jeremy Diaz. 2019. “Application of UNet Fully Convolutional Neural Network to Impervious Surface Segmentation in Urban Environment from High Resolution Satellite Imagery.” In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, 3915–18. IEEE.

McGranahan, Gordon, Deborah Balk, and Bridget Anderson. 2007. “The Rising Tide: Assessing the Risks of Climate Change and Human Settlements in Low Elevation Coastal Zones.” *Environment and Urbanization* 19 (1): 17–37.

Metcalfe, Daniel B, Thirze DG Hermans, Jenny Ahlstrand, Michael Becker, Martin Berggren, Robert G Björk, Mats P Björkman, et al. 2018. “Patchy Field Sampling Biases Understanding of Climate Change Impacts Across the Arctic.” *Nature Ecology & Evolution* 2 (9): 1443–48.

Mlambo, Reason, Iain H Woodhouse, France Gerard, and Karen Anderson. 2017. “Structure from Motion (SfM) Photogrammetry with Drone Data: A Low Cost Method for Monitoring Greenhouse Gas Emissions from Forests in Developing Countries.” *Forests* 8 (3): 68.

Munafò, Marcus R, Brian A Nosek, Dorothy VM Bishop, Katherine S Button, Christopher D Chambers, Nathalie Percie Du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J Ware, and John PA Ioannidis. 2017. “A Manifesto for Reproducible Science.” *Nature Human Behaviour* 1 (1): 1–9.

Musiega, Douglas E, Kazadi Sanga-Ngoie, and Kaoru Fukuyama. 2006. “A Framework for Predicting and Visualizing the East African Wildebeest Migration-Route Patterns in Variable Climatic Conditions Using Geographic Information System and Remote Sensing.” *Ecological Research* 21 (4): 530–43.

Olah, Chris, and Shan Carter. 2017. “Research Debt.” *Distill* 2 (3): e5.

Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. “You Only Look Once: Unified, Real-Time Object Detection.” In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779–88.

Rocklin, Matthew. 2015. “Dask: Parallel Computation with Blocked Algorithms and Task Scheduling.” In *Proceedings of the 14th Python in Science Conference*. Vol. 126. Citeseer.

Schmitt, Michael, and Xiao Xiang Zhu. 2016. “Data Fusion and Remote Sensing: An Ever-Growing Relationship.” *IEEE Geoscience and Remote Sensing Magazine* 4 (4): 6–23.

Shaman, Jeffrey, Susan Solomon, Rita R Colwell, and Christopher B Field. 2013. “Fostering Advances in Interdisciplinary Climate Science.” *Proceedings of the National Academy of Sciences* 110 (Supplement 1): 3653–56.

Simard, Marc, Keqi Zhang, Victor H Rivera-Monroy, Michael S Ross, Pablo L Ruiz, Edward Castañeda-Moya, Robert R Twilley, and Ernesto Rodriguez. 2006. “Mapping Height and Biomass of Mangrove Forests in Everglades National Park with SRTM Elevation Data.” *Photogrammetric Engineering & Remote Sensing* 72 (3): 299–311.

Solvik, Kylen, Anne M Bartuszevige, Meghan Bogaerts, and Maxwell B Joseph. 2020. “Predicting Playa Inundation Using a Long Short-Term Memory Neural Network.” *arXiv Preprint arXiv:2010.08605*.

Surazakov, Arzhan B, and Vladimir B Aizen. 2006. “Estimating Volume Change of Mountain Glaciers Using SRTM and Map-Based Topographic Data.” *IEEE Transactions on Geoscience and Remote Sensing* 44 (10): 2991–95.

Tayi, Giri Kumar, and Donald P Ballou. 1998. “Examining Data Quality.” *Communications of the ACM* 41 (2): 54–57.

Thatcher, Cindy A, Vicki Lukas, and Jason M Stoker. 2020. “The 3d Elevation Program and Energy for the Nation.” US Geological Survey.

Wallace, Luke, Arko Lucieer, Zbyněk Malenovskỳ, Darren Turner, and Petr Vopěnka. 2016. “Assessment of Forest Structure Using Two UAV Techniques: A Comparison of Airborne Laser Scanning and Structure from Motion (SfM) Point Clouds.” *Forests* 7 (3): 62.

Wang, Ning. 2019. “‘A Success Story That Can Be Sold?’: A Case Study of Humanitarian Use of Drones.” In *2019 IEEE International Symposium on Technology and Society (ISTAS)*, 1–6. IEEE.

Westoby, Matthew J, James Brasington, Niel F Glasser, Michael J Hambrey, and Jennifer M Reynolds. 2012. “‘Structure-from-Motion’photogrammetry: A Low-Cost, Effective Tool for Geoscience Applications.” *Geomorphology* 179: 300–314.

Wilkinson, Mark D, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3 (1): 1–9.

Woods, Andrew W, Rick E Holasek, and Stephen Self. 1995. “Wind-Driven Dispersal of Volcanic Ash Plumes and Its Control on the Thermal Structure of the Plume-Top.” *Bulletin of Volcanology* 57 (5): 283–92.

Wyder, Philippe Martin, Yan-Song Chen, Adrian J Lasrado, Rafael J Pelles, Robert Kwiatkowski, Edith OA Comas, Richard Kennedy, et al. 2019. “Autonomous Drone Hunter Operating by Deep Learning and All-Onboard Computations in GPS-Denied Environments.” *PloS One* 14 (11): e0225092.

Wyngaard, Jane, Lindsay Barbieri, Andrea Thomer, Josip Adams, Don Sullivan, Christopher Crosby, Cynthia Parr, Jens Klump, Sudhir Raj Shrestha, and Tom Bell. 2019. “Emergent Challenges for Science sUAS Data Management: Fairness Through Community Engagement and Best Practices Development.” *Remote Sensing* 11 (15): 1797.

Zhang, Wuming, Jianbo Qi, Peng Wan, Hongtao Wang, Donghui Xie, Xiaoyan Wang, and Guangjian Yan. 2016. “An Easy-to-Use Airborne LiDAR Data Filtering Method Based on Cloth Simulation.” *Remote Sensing* 8 (6): 501.

Zipkin, Elise F, Erin R Zylstra, Alexander D Wright, Sarah P Saunders, Andrew O Finley, Michael C Dietze, Malcolm S Itter, and Morgan W Tingley. 2021. “Addressing Data Integration Challenges to Link Ecological Processes Across Scales.” *Frontiers in Ecology and the Environment* 19 (1): 30–38.

# Figure Captions

Figure 1. Different kinds of resolution, with examples of lower and higher resolution data. Spatial resolution relates to pixel size, temporal resolution to observation frequency, radiometric resolution to the number of unique values, and spectral resolution to binwidth in the electromagnetic spectrum.